Outer membrane translocation of bulky small molecules by passive diffusion

Bert van den Berg[1#], Satya Prathyusha Bhamidimarri[2], Jigneshkumar Dahyabhai Prajapati[2], Ulrich Kleinekathöfer[2], Mathias Winterhalter[2]

[1]Institute for Cellular and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

[2]School of Engineering and Science, Jacobs University Bremen, 28759 Bremen, Germany

[#]To whom correspondence should be addressed. e-mail: bert.van-den-berg@ncl.ac.uk

Classification: BIOLOGICAL SCIENCES: Biophysics and Computational Biology

Keywords: CymA, outer membrane channel, cyclodextrin, passive diffusion, ligand-gating

Significance

The outer membrane (OM) of Gram-negative bacteria forms a protective layer on the outside of the cell that prevents unrestricted access of harmful compounds. For the acquisition of ions and nutrients, the OM contains two types of transport proteins: passive diffusion channels and active transporters. Due to the limited diameters of passive diffusion channels, bulky molecules such as iron-siderophores and complex oligosaccharides are assumed to be taken up exclusively by active transporters. Here we assert that this assumption is incorrect. Using a combination of biophysical and computational approaches we show that the OM protein CymA from *Klebsiella oxytoca* represents a novel paradigm in OM transport by mediating the passive diffusion of cyclic oligosaccharides (cyclodextrins) with diameters of ~15 Å.

Abstract

The outer membrane (OM) of Gram-negative bacteria forms a protective layer around the cell that serves as a permeability barrier to prevent unrestricted access of noxious substances. The permeability barrier of the OM results partly from the limited pore diameters of OM diffusion channels. As a consequence there is an OM "size-exclusion limit", and the uptake of bulky molecules with molecular masses more than ~600 Da is thought to be mediated by TonB-dependent, active transporters. Intriguingly, the OM protein CymA from *Klebsiella oxytoca* does not depend on TonB but nevertheless mediates efficient OM passage of cyclodextrins with diameters of up to ~15 Å. Here we show, by using X-ray crystallography, molecular dynamics simulations and single channel electrophysiology that CymA forms a monomeric 14-stranded β-barrel with a large pore that is occluded on the periplasmic side by the N-terminal 15 residues of the protein. Representing a novel paradigm in OM transport, CymA mediates the passive diffusion of bulky molecules via an elegant transport mechanism in which a mobile element formed by the N-terminus acts as a ligand-expelled gate to preserve the permeability barrier of the OM.

Introduction

The outer membrane (OM) of Gram-negative bacteria serves as an efficient permeability barrier for water-soluble and hydrophobic molecules. To obtain the necessary compounds for cell growth and function, the OM contains various β-barrel membrane proteins that serve as diffusion channels (1). Two classes of OM diffusion channels can be distinguished: non-specific porins and substrate specific channels. Due to the need to prevent unrestricted access of potentially toxic molecules, OM diffusion channels from both classes have pores that are relatively narrow with diameters of at most 7-8 Å. Because of these limited pore sizes, cellular entry by diffusion is thought to be effectively prevented for molecules larger than ~600 Da. However, many physiologically important molecules are larger than this "OM size-exclusion limit". Such compounds (*e.g.* iron-siderophores, vitamin B_{12} and complex oligosaccharides) are taken up by active OM transport proteins termed TonB-dependent transporters (TBDTs) (2). Although TBDTs have large barrels of 22 β-strands they do not have a permanently open pore due to a N-terminal "plug" or "cork" domain that completely fills the lumen of the barrel. The plug has a short sequence termed the TonB box which, upon

substrate binding by the transporter, interacts with a periplasmic domain of the TonB protein. TonB is part of the ExbBD-TonB inner membrane protein complex which functions as a proton pump to induce conformational changes in TonB as a result of the movement of protons. These conformational changes may then result in partial unfolding or ejection of the TBDT plug, resulting in a transient large channel through which the substrate passes into the periplasmic space. Thus, transport of bulky small molecules by TBDTs is an energy-dependent process that requires the proton motive force (PMF) across the inner membrane (2).

The OM protein CymA of *Klebsiella oxytoca,* a close relative of *K. pneumoniae,* is part of the *cym* (cyclodextrin metabolism) operon dedicated to the utilisation of cyclodextrins (CDs), cylindrical oligosaccharides containing six (α-CD), seven (β-CD) or eight (γ-CD) glucose units. Orthologs of CymA are present in the Enterobacteriaceae and Vibrionaceae (including *V. cholerae*), but only the protein from *K. oxytoca* has been studied. CDs are bulky molecules, with outer diameters of 13.7 Å for α-CD (MW 973 Da) and 15.3 Å for β-CD (MW 1135 Da), *i.e*. they are clearly too large to pass through known OM diffusion channels. CDs are formed extracellularly from starch by the action of secreted cyclodextrin-glucanotransferases (*cgt*). Besides *cymA* and *cgt*, the *cym* operon contains *cymE*, coding for a periplamic binding protein, *cymDGF* encoding an ABC transporter and *cymH* encoding a cytoplasmic cyclodextrinase that converts CDs into linear malto-oligosaccharides that then enter the maltose degradation pathway (Fig. 1; ref. 3). *E. coli* transformed with the *cym* operon grows on α-CD and β-CD as sole sources of carbon (4). In the absence of the *cymA* gene, growth on CDs is lost (4). In addition, CymA confers to an *E. coli* Δ*lamB* strain the ability to utilise maltopentaose and maltohexaose for growth (5). These data indicate that CymA functions as an efficient OM uptake channel for α- and β-CD as well as for linear malto-oligosaccharides. The relatively small size of CymA (~39 kDa) and sequence analysis suggest that CymA is not a TBDT but a regular diffusion channel, raising the question how such large substrates are transported across the OM without the requirement for an external energy source, and without compromising the OM permeability barrier.

To answer these questions we report here the X-ray crystal structures of CymA in the absence and presence of the substrates α- and β-CD. CymA forms a 14-stranded β-barrel

with a simple architecture for the extracellular loops. The N-terminal ~15 residues fold into the lumen from the periplasmic side and constrict the channel. The CymA-CD co-crystal structures reveal two α-CD binding sites inside the channel; one at the extracellular mouth and the other one near the periplasmic exit, delineating the passageway through the OM. The structures, together with molecular dynamics simulations and single-channel electrophysiology experiments, reveal a novel and elegant mechanism for the passive diffusion of bulky molecules while at the same time preserving the permeability barrier of the OM.

Results

The CymA N-terminus forms a mobile element in the channel lumen**.** After unsuccessful molecular replacement trials with crystals of the native protein, the structure of CymA was solved with phases obtained from a single anomalous dispersion (SAD) experiment on a selenomethionine (SeMet) substituted crystal. The CymA structure shows a monomeric 14-stranded β-barrel with relatively short extracellular loops (Fig. 2). A DALI search identified OmpG as the closest structural homologue in the database (PDB ID 2IWV; Z-score = 24, r.m.s.d. 2.9 Å; refs. 6,7). There is clear helical density inside the barrel lumen on the periplasmic side, resembling a post-cleavage autotransporter structure (8). Surprisingly, the helical density corresponds to the entire 15 residue-long cloning region including the hepta-histidine sequence (Methods) and the first two residues of CymA. Electron density for residues 3-10 is absent, presumably due to disorder, but the intervening distance of ~15 Å could easily be bridged by those missing eight residues. To our knowledge, this is the first example of a histidine tag forming part of a well-defined α-helix, and is a striking example of how environmental context shapes protein structure.

In addition to the SeMet structure we also solved a structure for the unlabeled ("native") His-tagged protein using data to 1.9 Å. Interestingly this structure, obtained using different crystallisation conditions relative to the SeMet protein, does not show any density for both the cloning region and the first 10 amino acids of CymA (Fig. 2), indicating that the N-terminus has likely moved out of the barrel. For both structures, none of the extracellular loops fold inwards to constrict the channel, contrasting with virtually all other OM diffusion channels (1). Consequently, the lumen of the barrel of the high-resolution structure is empty and forms a

circular channel with a fairly uniform diameter of ~11-14 Å that is filled with water and has a length of ~45 Å. When compared to *E. coli* OmpF it is clear that the CymA channel is much wider, with dimensions that appear a close match to that of α-CD (Fig. S1). Taken together, the structural data for the tagged protein suggest that the N-terminus of CymA is mobile and might be able to move in and out of the large-diameter channel.

We next asked where the N-terminus is located in the wild type, tag-less protein. To answer this question we purified and determined the crystal structure of non-tagged CymA (termed "wild type"). As shown in Fig 2, the N-terminus of wild type CymA is present inside the barrel. The side chain of Arg5 interacts with three glutamic acid residues in the barrel wall (Glu30, Glu287 and Glu289), and is likely to be important for positioning of the N-terminus given that few other interactions are evident. Interestingly, residue Arg -11 of the cloning tag makes very similar interactions in the structure of the SeMet protein (Fig. S2). In the wild type protein structure, density is missing for residues 10-21 but, similar to the his-tagged protein, the intervening distance (~17 Å) can be bridged by those missing residues. Both His-tagged CymA and wild type CymA migrate at their expected molecular weight in SDS-PAGE gels, arguing against proteolytic cleavage in the N-terminus as a source of the disorder. To obtain a more complete picture of the wild type protein we modeled the missing residues (Glu10-Phe21) followed by 100 ns of equilibrium molecular dynamics (MD) simulation (Fig. 3). The simulation suggests that the structure is stable (Fig. S3) and confirms the likely importance of Arg5 for positioning of the N-terminus inside the barrel. Moreover, high r.m.s.f. values and the presence of relatively few interactions between Glu10-Phe21 and residues in the barrel wall provide an explanation for the fact that this segment is not observed in the crystal structure (Fig. 3). The wild type CymA model shows that the presence of the N-terminus inside the barrel restricts the diameter of the channel dramatically (Fig. 3), and this has important implications for substrate transport.

To provide supporting evidence for the structural data we performed single channel electrophysiology on CymA proteins. Wild type CymA forms ion-permeable channels with an average conductance of 1.16 nS in solvent free membranes (Fig. 4). The conductance states and the open probabilities are dependent on the magnitude and polarity of the applied voltage. At low voltages, the channel is predominantly in a non-conducting state (Fig. 4A).

Since OM Donnan potentials are low to non-existent under physiological conditions (< 30 mV; ref. 9) the CymA channel is likely to be mostly closed *in vivo*. With increasing positive voltages, CymA adopts a stable open state whereas at negative voltages the channel displays frequent channel closures even at high potential differences (Fig. 4). Taken together with the crystal structures we propose that the open state results from a voltage-induced displacement of the N-terminus from the lumen of the barrel. To confirm this notion we constructed and purified (see Methods) an N-terminal truncation mutant which has the first 15 residues removed (ΔCymA). Single channel experiments demonstrate that the mutant channel is completely open at positive and negative voltages (Fig. 4E). Moreover, ΔCymA is silent, indicating that the extracellular loops are rigid and do not fold inwards to transiently close the channel. These results confirm previous electrophysiology data, including the suggestion that the channel "contains a section, which is mobile and can block the channel at least partially" (9). Our data show that this assumption was correct and that the mobile element corresponds to the N-terminus of CymA moving into and out of the barrel lumen. The final protein that we tested was the R5A site-directed mutant. Based on the crystal structure of WT CymA (Fig. S2 and Fig. 3B), Arg5 is likely to play an important role in constraining the N-terminus inside the lumen of the barrel by interacting with several side chain carboxyl groups in the barrel wall. The substitution of Arg5 with alanine should abolish those interactions. The electrophysiology data for the R5A mutant protein indeed show that even at low voltages the channel is open almost permanently (Fig. 4F), confirming the important role of Arg5 in constraining the N-terminus.

CymA contains two binding sites for cyclodextrins. We next sought to characterize the interaction of CD substrates with the CymA channel. For this, we soaked α- and β-CD into pre-grown his-tagged CymA crystals that diffracted to high resolution. The structure for the complex of CymA and α-CD (1.7 Å resolution) does not show any density for the N-terminus and the cloning region. Instead, well-defined density is present inside the lumen of the barrel, indicating two binding sites for the cyclic oligosaccharide (Fig. 5 and Fig. S4). One site is formed by the extracellular mouth of the channel and likely represents the initial CD binding site. We will refer to this site as the "entry site". The second binding site is ~30 Å away and at a position close to where the centre of the OM bilayer would be (Fig. 5). Beyond this position the channel widens towards the periplasmic space and we will therefore designate the

second binding site the "exit site". Strikingly, the orientation of the α-CD molecule is completely different in both sites. In the entry site, the plane of the CD ring is approximately parallel to the membrane surface. By contrast, the CD ring in the exit site has rotated almost 90 degrees and is now almost perpendicular to the membrane plane (Fig. 5). Thus, during diffusion through the channel the substrate molecule reorients itself.

An additional interesting feature of the CymA-CD co-crystal structures is the presence of tubular density inside the CD cavities (Fig. S4). These densities most likely correspond to partially ordered C_8E_4 molecules used for purification and crystallisation, and demonstrate the well-established property of CDs to form complexes with hydrophobic compounds (11). It may therefore be possible to utilise CymA as a means of delivery of useful hydrophobic molecules in synthetic biology or biotechnological applications involving Gram-negative bacteria.

CD binding to CymA results in only minor conformational changes. Given the large size of the CD molecules it might be expected that substantial conformational changes in CymA would be required for sugar binding. However, the structures show that for both binding sites the structural changes in the channel upon substrate binding are very small. Within the entry site, just three amino acid side chains undergo conformational changes more than 1 Å: Asn40 in loop L1, Glu80 in loop L2 and Trp303 in loop L7 (Fig. S5A). The latter side chain reorients itself to form stacking interactions of its aromatic ring with a sugar moiety, as commonly observed in sugar binding and transport proteins (12,13). Additional stacking interactions with the entry site-bound CD are mediated by Tyr84 and Trp169. In addition to those interactions, numerous hydrogen bonds are present between the entry site CD and polar and charged side chains of CymA. Compared to the entry site, the exit site α-CD molecule makes fewer interactions with residues in the barrel wall due to its orientation and off-center position inside the channel. The structural changes in the exit site upon substrate binding are also very small and restricted to the side chains of residues E113, E275 and Q319. Only one aromatic residue (Tyr154) makes stacking interactions with sugar moieties (Fig. S5B).

CD binding by CymA is unidirectional. The CymA structure complexed to α-CD suggests

that the extracellular mouth of the channel has evolved to bind the substrate with high affinity. We next asked if we could obtain evidence for directionality of substrate binding by single channel electrophysiology. When substrate is added to wild type CymA from the *cis* side, no current blockages are observed. By contrast, addition of substrate from the *trans* side blocks the channel very efficiently with a stability constant of 34,000 ± 2700 M^{-1} (Fig. 6C), confirming the high affinity of CymA for α-CD from previous multichannel titrations (10) and suggesting that the *trans* side corresponds to the extracellular milieu. Even though the channel is fully open at the applied voltage (+100 mV; Fig. 4), it could be argued that the presence of the mobile N-terminus could hamper substrate binding from the periplasmic side. To test this notion we also performed substrate addition experiments for ΔCymA. Remarkably, the dependence of substrate binding on the orientation of the channel is also observed for the N-terminal deletion mutant, with efficient channel blockage only observed from the *trans* side (Fig. 6; stability constant 35,000 ± 1300 M^{-1}). These observations indicate that the substrate does not, or only very inefficiently, enter the channel from the periplasmic space, which in turn suggests that the exit site CD molecule observed in our co-crystal structure has diffused there from the entry site. Furthermore, the affinities of α-CD for the N-terminal deletion variant and the wild type protein are very similar, indicating that the N-terminus does not play a role in substrate binding.

The closely matching dimensions of the CymA channel and α-CD (Fig. 5) raise the question how the larger β-CD molecule traverses the channel, and we therefore also determined a co-crystal structure of CymA with β-CD. As for α-CD, electron density for a large cyclic oligosaccharide is present in two sites. Surprisingly, the density in the entry site corresponds to α-CD whereas the density in the exit site can be assigned unambiguously to β-CD (Fig. 5 and Fig. S4). The β-CD molecule has a similar orientation as the exit site α-CD, with the plane of the ring being approximately perpendicular to the OM plane. While the orientation is similar, the β-CD molecule is shifted ~4-5 Å towards the periplasmic space compared to α-CD and also occupies a more central position inside the CymA channel (Fig. S4C). As is the case for α-CD, the binding of β-CD introduces only small conformational changes in a few residues lining the channel (Fig. S5C). The fact that α-CD is bound in the entry site must be due to a contamination in the β-CD preparation coupled to a much higher affinity of α-CD for

the entry site relative to β-CD. Indeed, single channel measurements clearly show the lower affinity of β-CD for CymA (Fig. 6D), with a stability constant of 1600 ± 200 M-1 compared to 34000 ± 2700 for α-CD. We hypothesise that the size of β-CD makes it unlikely to bind in the entry site in a similar way as α-CD, *i.e*. with the plane of the ring parallel to the OM plane. In addition, the CymA channel is likely too narrow for transport of γ-CD (17 Å outer diameter), which could explain why *K. oxytoca* does not grow on this compound (3). It appears that CymA has evolved for transport of α-CD, and it would therefore be interesting to determine whether α-CD predominates as a carbon source *in vivo*.

A dynamic view of CD transport. To complement the crystallographic snapshots we performed unbiased MD simulations of α-CD bound to the entry and exit sites of CymA, starting from the crystal structure coordinates. For the entry site we observe stable binding of the α-CD molecule only when the three acidic residues (Glu80, Glu85 and Glu205) that contact the sugar are protonated. When those residues are charged, the α-CD molecule is clearly unstable and assumes a wide range of orientations during the simulation (Fig. S6). The calculated pKa values for the glutamic acid residues are elevated by ~1 pH unit in the presence of the bound substrate, but are still well below the crystallisation and soaking pH of 7.5 (Methods). Further studies are required to shed light on the possible role of binding site protonation in substrate binding and release. By contrast, the α-CD molecule in the exit site is stably bound for the duration of the simulation regardless of the protonation states of the two acidic molecules (Glu113 and Glu275) that contact the substrate.

We next investigated whether the substrate reorientation during transport, inferred from the crystal structures, could be recapitulated in a steered MD simulation. To avoid the complication of the presence of the N-terminus within the lumen of the barrel we used the coordinates of the co-crystal structure of CymA with α-CD (PDB ID 4D5B), *i.e*. the N-terminus is absent during the simulation. The α-CD molecule was placed approximately 40 Å from the extracellular mouth of the pore and was pulled with constant speed towards the periplasmic space (see Methods). The results of the simulations indeed match the co-crystal structures, though we emphasize that the steered MD simulations should only be viewed as a qualitative indication of the translocation pathway. The substrate binds to the entry site with the plane of

the CD ring almost parallel to the membrane place (~160° angle). Upon release from the entry site the substrate tilts immediately (~90°angle) and remains in this orientation until it reaches the periplasmic space (Fig. 7 and Supplementary Movie). We speculate that the rotation of the cyclic sugar during diffusion serves to lower its affinity for the channel, which would prevent substrate from stalling prior to exiting into the periplasmic space.

We also carried out steered MD simulations for β-CD to investigate whether this compound can translocate through the CymA channel, as implied by growth data (4). The force profiles suggest that while the average forces required for pulling β-CD through the channel are somewhat higher than those for α-CD, the maximum forces are similar (8-10 kcal/mole/Å) (Fig. S7). This suggests that β-CD can traverse the channel but that permeation is likely to be less efficient compared to α-CD. In an interesting contrast with α-CD, β-CD appears to enter and pass through CymA close to its orientation in the crystal, with the plane of the ring approximately perpendicular to the OM plane (Fig. 7). This result reinforces our notion that β-CD does not bind to the entry site.

Discussion

A model for OM diffusion of bulky small molecules. The most interesting part of the CymA structure is the N-terminus, which forms a mobile element that can move in and out of the channel lumen. What is the role of the N-terminus, given that our data indicate it is dispensable for substrate binding (Fig. 6)? By constricting the channel, the N-terminal ~15 residues likely preserve the permeability of the OM, which might otherwise be compromised. Constriction elements composed of extracellular loops that fold inwards to decrease the effective diameter of the barrel are found in OM channels composed of 16 or more β-strands such as OmpF (Fig. S1) (1). Those loops tend to be very stable because they make many interactions with the interior surfaces of the barrel or with other loops. In the case of CymA a relatively short segment (the N-terminus) inserts into the channel from the periplasmic side and interacts with the barrel wall. To gain insight into the energetics of this interaction we performed a one-dimensional potential of mean force (PMF) calculation of the N-terminus in wild type CymA using metadynamics (14). The result shows that the position of the N-terminus in the crystal corresponds to the free energy minimum (Fig. S8). Moreover,

movement of the N-terminus into the periplasmic space requires passing several energy barriers. Thus, in order to generate an open channel the interactions of the N-terminus with the barrel need to be disturbed. Considering the fact that only a limited number of N-terminal residues interact with the barrel wall (Fig. 3B), we propose that the CD substrate moving down from the entry site could provide the driving force for channel opening, possibly by engaging the acidic residues of the channel wall that interact with the important residue Arg5 (Fig. 8). CymA can therefore be considered a ligand-gated OM diffusion channel. Future experiments and more detailed molecular dynamics simulations will be required to obtain insights into the displacement mechanism of the N-terminus by CDs and whether other molecules can open the channel.

OM diffusion channels that allow permeation of molecules larger than the "OM size exclusion limit" have been identified previously. In all cases, however, the substrates have structures allowing them to pass the channel in a linear fashion, *e.g*. malto-oligosaccharide passage through *E. coli* LamB (12). Our study now shows that bulky molecules such as CDs can also enter cells efficiently through OM diffusion channels. This then poses the question as to why cells spend PMF-derived energy to take up bulky substrates via TonB-dependent transporters (TBDTs). As CymA illustrates, substrate size is not the decisive factor that necessitates active transport. Rather, it is likely the need to bind substrate very tightly because it is scarce and/or valuable. The external energy input is required to generate the large conformational changes necessary for substrate release. In the case of TBDTs, substrate release is coupled to formation of a transient channel into the periplasmic space. By contrast, the mobile N-terminal constriction domain of CymA is likely displaced by the incoming substrate, providing an elegant way for cells to take up bulky substrates without compromising the permeability barrier of the OM.

Methods

Cloning, overexpression and purification of His-tagged CymA. The mature part of the *cymA* gene from *K. oxytoca* was synthesised (MWG Genomics) and cloned into the pB22 expression vector under control of the arabinose promoter. After cleavage by signal peptidase, the N-terminal sequence of the protein is as follows: (-

14)ANVRLQHHHHHHHHLE(0)-CymA. Protein expression was performed in C43 Δ*cyoABCD* cells by growing in LB medium till OD_{600} ~0.6 at 37 °C, followed by induction with 0.1% arabinose for 16 hrs at 20 °C (final OD ~1.2-1.5). Cells were harvested by centrifugation, resuspended in TSB buffer (20 mM Tris/300 mM NaCl pH 8) and broken via one pass through a cell disrupter operated at 23 kpsi (Constant Biosystems). Total membranes and cell debris were collected by ultracentrifugation at 42,000 rpm for 45 mins (45Ti rotor, Beckman). Membranes were homogenised with 3% (w/v) Elugent (Calbiochem) in TSB (100 ml for 6 liters of culture) followed by stirring at 4 °C O/N. After ultracentrifugation (30 mins 42 krpm; 45Ti rotor), the supernatant was loaded onto a 10 ml nickel column equilibrated in TSB + 0.2% LDAO (chelating sepharose; GE Healthcare). The column was washed with 15 CV buffer with 25 mM imidazole and the protein was eluted with 3 CV buffer + 250 mM imidazole. After concentration (50 kDa cutoff; Millipore) the protein was applied to a Superdex-200 16/60 gel filtration column (GE Healthcare) equilibrated in 10 mM Hepes/100 mM NaCl/0.05% LDAO pH 7.5. For polishing and buffer exchange prior to crystallisation, a second gel filtration column was run in 10 mM Hepes/100 mM LiCl/0.35-0.4% C8E4 pH 7.5. The protein was concentrated to 10-15 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C. The yield of purified CymA is ~ 4 mg per 6 liters of rich medium. For production of selenomethionine-substituted protein, C43 Δ*cyoABCD* cells were grown at 37°C in LeMasters/Richards (LR) minimal medium with 0.25% glycerol as the carbon source, using the methionine biosynthesis inhibition method (15). Approximately 20 mins before induction, amino acids were added (K/T/F; 100 mg/l; L/I/V 50 mg/l) as well as SeMet (60 mg/l). Induction was carried out by adding 0.5% arabinose, followed by 16 hrs growth at 30 °C. The final OD_{600} was ~ 1.0. The SeMet-substituted protein was purified as described above for the native protein; in this case, the yield was ~ 2.5 mg/10 liter culture. For production of non-His tagged (wild type) CymA, the porin-deficient *E. coli* Bl21 omp8 strain was transformed with the pUC18-derived pCYMA plasmid which expresses CymA constitutively (5). Four liters of cells were grown in LB medium at 37 °C throughout; curiously, growth at lower temperatures did not result in any expression. For CymA purification, the total membrane pellet was extracted twice with 100 ml 0.5% sarkosyl in 20 mM Hepes, pH 7.5 to selectively remove inner membrane proteins. The OM-enriched pellet was extracted with 1% LDAO in 10 mM Hepes/50 mM NaCl pH 7.5 by homogenisation and stirring for 2 hrs at 4 °C. After ultracentrifugation, the supernatant was loaded onto a 10 ml Q-Sepharose fast flow columns equilibrated in 0.2% LDAO, washed with

LDAO buffer containing 100 mM NaCl, and eluted with 500 mM NaCl. After concentration the protein was purified on Superdex-200 26/60 in 10 mM Tris/100 mM NaCl/0.05% LDAO pH 8, followed by a linear salt gradient on Resource Q (6 ml) at pH 8. Appropriate fractions were pooled and run on a final Superdex S-200 16/60 gel filtration column in 10 mM Hepes/100 mM LiCl/0.35-0.4% C8E4 pH 7.5. The protein was concentrated to 8 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C. The yield of purified native CymA is ~2 mg per 4 liters of medium.

Construction, expression and purification of ΔCymA. pCYMA was amplified using Phusion DNA polymerase (Thermo Scientific) and the oligonucleotide primers FP (5'-GAAAGTTTTTTTTCGTTTGGTGGCCAT-3') and RP (5'-TGCAAATGAATGTACGGGCGCTGTGA-3'). The PCR product was digested with DpnI, phosphorylated using T4 PNK (Thermo Scientific) and ligated using T4 DNA ligase (Thermo Scientific). The ligation product was transformed into DH5α cells by electroporation. ΔCymA was expressed and purified as published previously for wild type CymA with slight modifications (5). Cells were grown at 37 °C overnight to an OD_{600} of 0.8-1.1 in LB medium, harvested by centrifugation and re-suspended in 50 mM Potassium phosphate with protease inhibitor (1mM PMSF) at pH 7.5 (Buffer A). 3-5 passages of French press were employed to lyse the cells at 16,000 psi. Intact cells were removed by centrifugation at 6000 x *g* for 1 hour. Ultracentrifugation was done at 100,000 x *g* for 90 mins to pellet cell envelopes. Buffer A with 0.5% sarcosyl was used to solubilize inner membrane components, stirred for 1 hour at room temperature followed by centrifugation at 100,000 x *g* for 90 mins. This pre-extraction procedure was repeated once. The pellet was washed in Buffer A and re-suspended in Buffer A with 100mg/ml lysozyme, 3 mM NaN_3 and the mixture was left for overnight stirring at 37 °C. The solution was centrifuged at 100,000 x *g* for 90 mins and the OM was solubilized by resuspending in Buffer A with 5mM EDTA and 5% octyl- polyoxyethylene (OPOE, Bachem Biochemica GmbH, Heidelberg, Germany) and stirred at room temperature for 1 hr. After centrifugation, the supernatant was dialyzed against 10 mM Potassium phosphate pH 7.5 containing 0.6 % OPOE (Buffer B) and loaded onto a Mono Q-HR 5/5 column equilibrated in Buffer B. ΔCymA was eluted with Buffer B containing 1 M NaCl using a linear gradient to 1 M NaCl), concentrated and flash-frozen for storage at -80 °C.

Single channel electrophysiology. Solvent-free membranes were made according to Montal and Muller (16). Briefly, a Teflon cuvette with two symmetric chambers partitioned by a 25 µm thick Teflon film with an aperture size of approximately 50-100 µm was used for reconstitution of protein. The aperture was impregnated with 1% hexadecane in n-hexane to make it more hydrophobic. The aqueous phase was buffered with 10 mM MES at pH 6. The membrane was formed using a solution of 5 mg/ml DPhPC in n-pentane. Ag/AgCl electrodes (World Precision Instruments, Sarasota, FL) were used to measure the electric current. One of the electrodes was connected to the ground (*cis* side of the membrane). The other electrode was connected to the head stage of an Axopatch 200B amplifier (Axon Instruments). Purified porin in detergent solution was added to the *cis* side of the membrane. The conductance measurements were performed by Axopatch 200B amplifier in the voltage clamp mode and digitized by Axon Digidata 1440A digitizer and controlled by Clampex software (Axon Instruments, Foster city, CA). The current traces were filtered by low pass Bessel filter at 10 kHz and recorded with a sampling frequency of 50 kHz. Data was analysed by the Clampfit program.

Crystallisation and structure determination. Initial vapour diffusion crystallisation trials were set up using a Mosquito crystallisation robot (TTP Labtech) with sitting drops at 20 °C with in-house screens as well as the commercial MemGold 1, MemGold 2 and Morpheus screens (Molecular Dimensions). If necessary, crystal optimisation was performed using larger-scale sitting drop or hanging drops. The SeMet crystal of his-tagged CymA used for data collection grew in space group $P2_1$ from Morpheus 1-35 (0.09 M NPS, 0.1 M buffer system 3 pH 8.5, 20% v/v glycerol, 10% PEG 4K). The high-resolution his-tagged CymA dataset grew from MemGold 1, 2-13 (0.1 M Tris pH 8.0, 0.3 M magnesium nitrate hexahydrate, 23% PEG 2K) (space group $P2_1$), whereas the tag-less wild type protein crystals were obtained from ~35% PEG400, 0.4 M sodium bromide, 50 mM MES pH 6 or 50 mM glycine pH 9 (space group $C222_1$). The his-tagged protein crystals used for CD soaking grew in Morpheus 1-30 (0.09 M NPS, 0.1 M buffer system 2 pH 7.5, 20% ethylene glycol, 10% PEG 8K (space group $P2_12_12$). From substrate-induced channel blockages in multi-channel electrophysiology experiments the binding constants were previously estimated to be ~30 µM for α-CD and 0.5 mM for β-CD (10). Based on these data, solutions of α- and β-CD (Sigma) were made in water at 100 mM and 50 mM respectively and diluted 10-fold into the crystallisation hit

solution containing 0.5% C_8E_4. This was then added in a 1:1 ratio to the crystal drop giving final concentrations of 5 mM and 2.5 mM (β-CD), and incubated for 48 hrs at 20 °C. Co-crystallisation experiments with α- and β-CD produced only small crystals.

Initially we attempted molecular replacement for the native high-resolution CymA dataset using the distantly related OmpG OM channel (~15% sequence identity to CymA) as a search model, but we were not successful. We then generated crystals of SeMet-substituted protein and collected three single anomalous dispersion (SAD) datasets at the peak wavelength on a single, bar-shaped crystal. The datasets were processed using XDS (17) or HKL2000 (18) and merged to obtain a single, high-redundancy dataset. All four selenium sites within the asymmetric unit were found using SOLVE (19), and RESOLVE AUTOBUILD was used to generate a partial model. The best-defined monomer of the AU was extended manually within Coot (20) and used as a search model for molecular replacement (MR) of the high-resolution native data using Phaser (21). A single, clear solution was found that was used for RESOLVE autobuilding, resulting in a model that was ~80% complete. Several rounds of Phenix refinement (22) and manual model extension were subsequently carried out to arrive at a final model with good statistics (R/R_{free} = 17.4/20.7%; data to 1.8 Å resolution). This model was used to solve the other CymA datasets via MR. Data collection and refinement statistics are summarised in Table S1.

Molecular Dynamics simulations. To study the structure and stability of the CymA channel and especially of the N-terminus inside the pore, CymA (PDB ID 4D51) was simulated in a POPC (1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine) bilayer consisting of 230 lipids per unit cell. The missing residues (Glu10-Phe21) of the CymA N-terminus were predicted using Modeller 9.12 (23). The system was solvated using the TIP3P water model and neutralized with 6 potassium ions leading to a total atom number of 91342. Subsequently, an energy minimization was performed followed by a 10 ns equilibration in an NVT ensemble at 300 K with position restraints on protein and lipids. In a further step, the position restrains were removed and the system was equilibrated for 100 ns at 1 bar. All MD simulations were performed using the GROMACS 4.6.5 package (24) using the standard CHARMM36 force field (25). The cutoff for short-range electrostatics and van der Waals interactions was set to 1.2 nm and long-range electrostatics was treated using the particle-mesh Ewald (PME)

method with grid size 0.1 nm. Moreover, all bonds were constrained using the LINCS method (26) to enable a time step of 2 fs.

In a second set of simulations, four different systems including α-CDs were built from the corresponding crystal structure (PDB ID 4D5B). To this end, the N-terminal residues (Ala1-Ala13) were assumed to be outside the channel unlike the system described above which has an inwardly oriented N-terminus. Two separate systems were built containing either an α-CD at the entry or at the exit binding site. Initially all the ionizable residues of the protein were kept in the charged state. However, the α-CD molecules were not very stably bound especially at the entry site. Therefore, we performed pKa calculations using PROPKA (27) in the absence and presence of α-CD. The output suggested that there is a change in the pKa values by about 1 pH unit in the presence of α-CD mainly for the three glutamate residues (Glu80, Glu85 and Glu205) at the entry site and two glutamate residues (Glu113 and Glu275) at the exit site. Based on these findings, two more systems were built with the above sets of the glutamate residues in their protonated state and the remaining ionizable residues unchanged. The equilibration procedure described above was followed by unbiased simulations of two times 50 ns for each of the four system variants.

In addition, constant velocity steered molecular dynamics (SMD) was performed to obtain an initial, qualitative idea of the translocation of α-CD through the pore. For this simulation the abovementioned five glutamate residues were protonated. Initially, the α-CD was placed outside the pore on the extracellular side and steered with a pulling velocity of 1Å/ns and a spring constant of 100 kJ/mol/nm^2.

Acknowledgements

We are indebted to Prof. Robert Gennis (University of Illinois) for his kind gift of the C43 Δ*cyoABCD* expression strain. We also thank Prof. Roland Benz (Jacobs University Bremen) for bringing the unusual properties of CymA to our attention. Jing Lu (Jacobs University Bremen) helped with the initial simulations. We would like to thank the staff of the Diamond Light Source beam lines i02 and i03 for beam time and assistance. The research leading to these results has received support from the Innovative Medicines Initiatives Joint Undertaking under Grant Agreement No. 115525, resources which are composed of financial contributions

from the European Union's seventh framework programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations companies in kind contribution. Furthermore, we acknowledge funding from ITN-2014-607694-Translocation.

Author contributions

BvdB cloned, purified and crystallised CymA, carried out structure determinations, designed research and wrote the manuscript. SPB carried out single channel electrophysiology experiments. JDP performed molecular dynamics simulations. UK and MW designed research and co-wrote the manuscript.

Accession codes

Coordinates and structure factors for apo his-tagged SeMet CymA, his-tagged native CymA, wild type CymA and the co-crystal structures of his-tagged CymA with α-CD and β-CD have been deposited in the Protein Data Bank with accession codes 4V3G, 4V3H, 4D51, 4D5B and 4D5D respectively.

References

1. Nikaido H (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* 67:593-656.
2. Noinaj N, Guillier M, Barnard TJ, Buchanan SK (2010) TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu Rev Microbiol* 64:43-60.
3. Pajatsch M, Gerhart M, Peist R, Horlacher R, Boos W, Böck A (1998) The periplasmic cyclodextrin binding protein CymE from Klebsiella oxytoca and its role in maltodextrin and cyclodextrin transport. *J Bacteriol* 180, 2630-2635.
4. Fiedler G, Pajatsch M, Böck A (1996) Genetics of a novel starch utilisation pathway present in Klebsiella oxytoca. *J Mol Biol*. 256, 279-291.
5. Pajatsch M, Andersen C, Mathes A, Böck A, Benz R, Engelhardt H (1999) Properties of a cyclodextrin-specific, unusual porin from Klebsiella oxytoca. *J Biol Chem*. 274, 25159-25166.

6. Yildiz O, Vinothkumar KR, Goswami P, Kuhlbrandt W (2006) Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J* 25, 3702-3713.
7. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucl. Acids Res*. 38, W545-549.
8. Grijpstra J, Arenas J, Rutten L, Tommassen J (2013) Autotransporter secretion: varying on a theme. *Res Microbiol*, 164, 562-582.
9. Sen S, Hellman J, Nikaido H (1988) Porin channels in intact cells of *Escherichia coli* are not affected by Donnan potentials across the outer membrane. *J. Biol Chem*. 263, 1182-1187.
10. Orlik F, Andersen, C, Danelon, C, Winterhalter, M, Pajatsch, M, Böck, A, Benz, R (2003) CymA of Klebsiella oxytoca outer membrane. Binding of cyclodextrins and study of the current noise of open channels. *Biophys J*. 85, 876-882.
11. Lakkakula JR and Macedo-Krause RW (2014) A vision for cyclodextrin nanoparticles in drug delivery systems and pharmaceutical applications. *Nanomedicine* 9, 877-894.
12. Schirmer T, Keller TA, Wang YF, Rosenbusch JP (1995) Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* 267, 512-514.
13. Sun L, Zeng X, Yan C, Sun X, Gong X, Rao, Y, Yan N (2012) Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature* 490, 361-366.
14. Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett*. 100, 020603.
15. Doublie S, Carter CW Jr (1992) Preparation of selenomethionyl protein crystals, in Ducruix A and Giegg R (eds) Crystallisation of nucleic acids and proteins. A practical approach. Oxford University Press.
16. Montal M, Mueller P (1972) Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl. Acad. Sci. USA* 69, 3561-3566.
17. Kabsch W (2010) XDS. *Act Crystallogr* D66, 125-132.
18. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
19. Terwilleger TC, Berendzen J (1997) Automated MAD and MIR structure solution. *Acta Crystallogr* D55, 849 – 851.

20. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr* D60, 2126 –2132
21. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr* D61, 458 – 464.
22. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr* D68, 352-367.
23. Sali A, Blundell TL (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* 234, 779–815.
24. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4, 435–447.
25. Huang J, Roux B (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* 34, 2135–2145.
26. Hess B, Bekker H, Berendsen HJC, Johannes GEMF (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18, 1463–1472.
27. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7, 525–537.
28. Smart OS, Neduvelil JG, Wang X, Wallace B, Sansom MS (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* 14, 354–360.
29. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun*. 185, 604-613.

Figure Legends

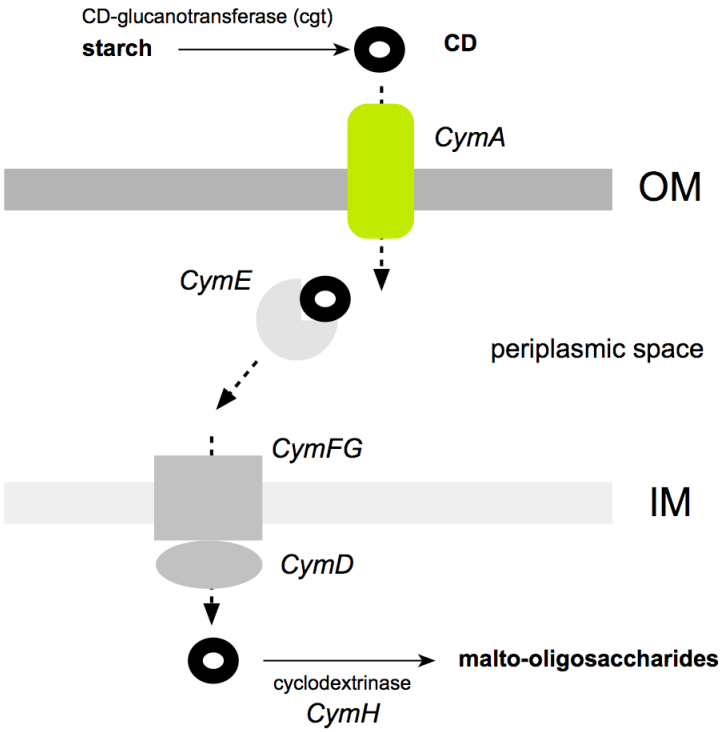

Figure 1 Schematic overview of cyclodextrin uptake and metabolism in *K. oxytoca*.

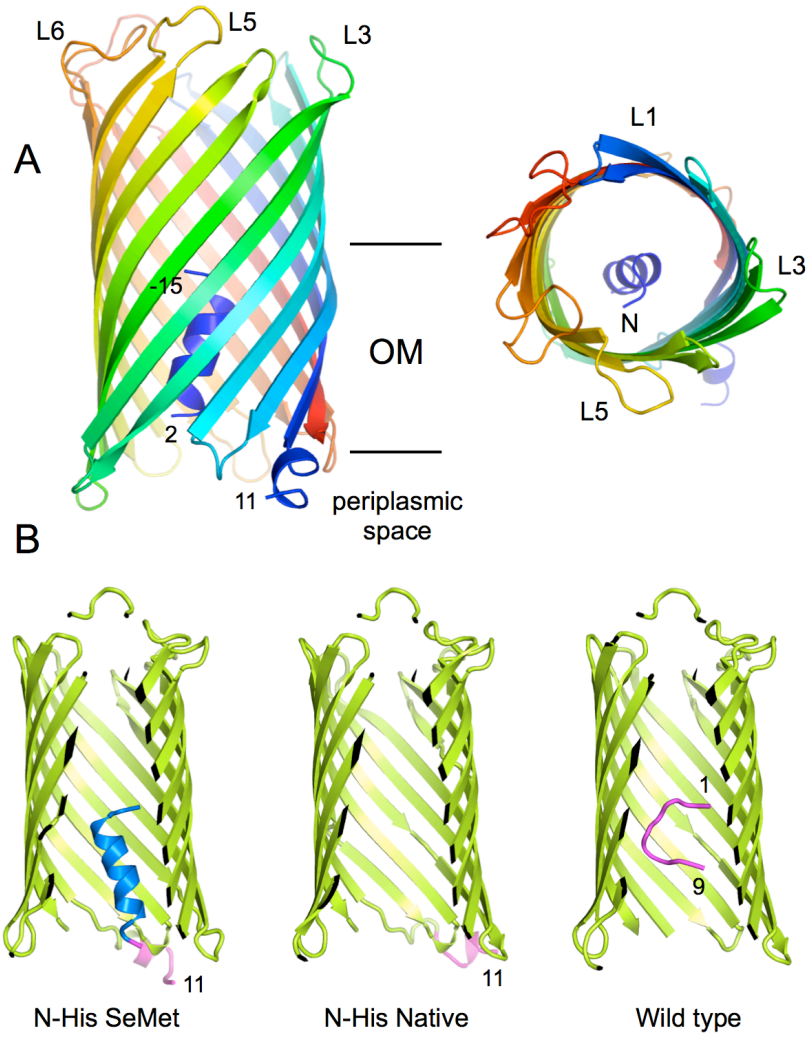

Figure 2 Overview of CymA crystal structures. (A) Cartoon model of SeMet CymA viewed from the side (left panel) and from the extracellular side in rainbow colouring. The numbering for the visible residues of the N-terminus (blue) is shown. Negative numbers correspond to residues of the cloning region. (B) Comparison of the interior of the barrel lumen for his-tagged SeMet CymA (PDB ID 4V3G), his-tagged CymA (PDB ID 4V3H) and wild type CymA (PDB ID 4D51). The cloning region is coloured blue and the first 17 residues of CymA are shown in magenta. Selected residues are numbered.

Figure 3

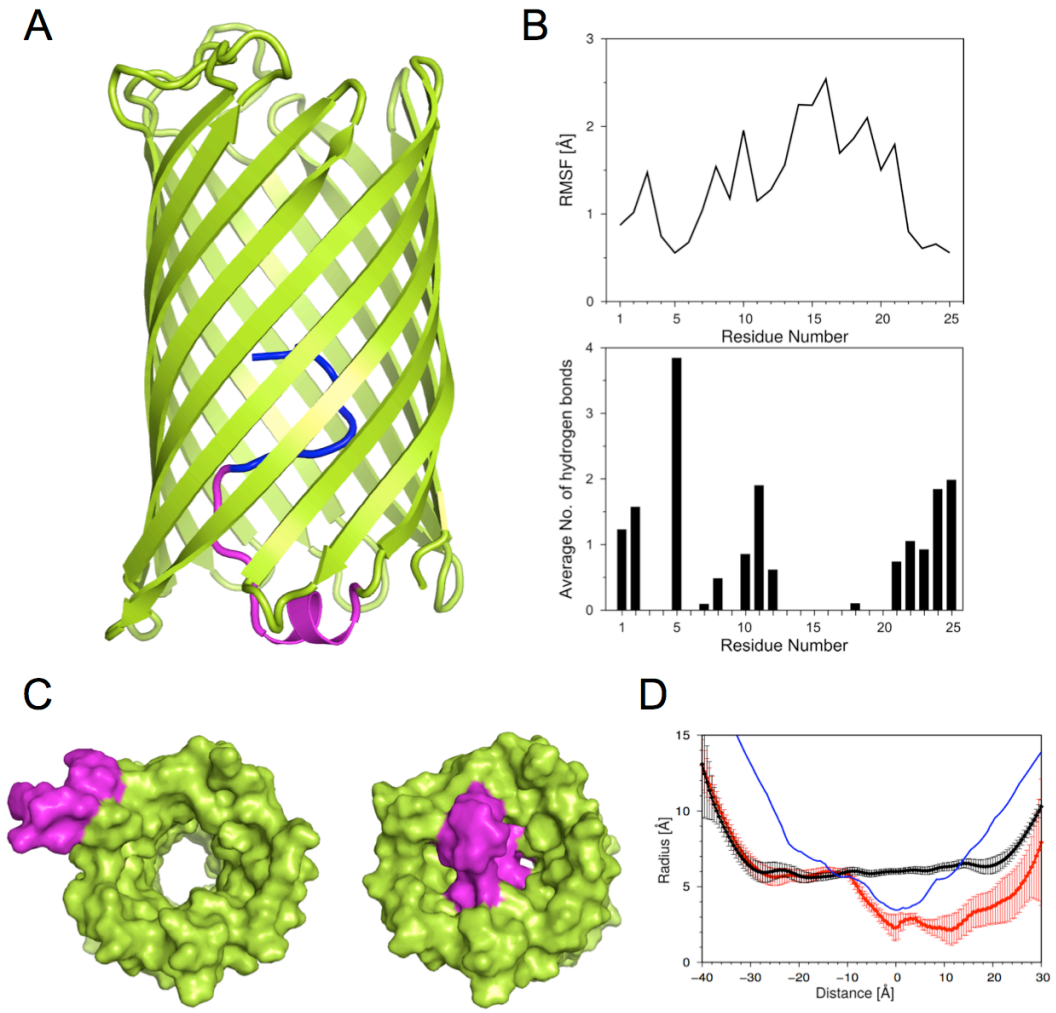

Figure 3 Constriction of the CymA channel by the N-terminus. (A) Cartoon of wild type CymA with residues Glu10-Phe21 (coloured magenta) modeled followed by a 100 ns MD simulation. Residues 1-9 visible in the crystal structure are shown in blue. (B) Root-mean-square-fluctuations (top panel) and hydrogen bonding interactions (bottom panel) of the N-terminus (residues 1-25) with residues of the barrel wall. (C) Surface views from the periplasmic space for CymA with the N-terminus expelled from the barrel (PDB ID 4V3H) and for the wild type CymA model (right panel), illustrating the size difference of the channel. (D) Average pore radii derived from unbiased MD simulations with the corresponding standard deviations for wild type CymA with (red) and without (black) the N-terminus bound inside the barrel. For comparison, the pore radius from the high-resolution crystal structure of OmpF (PDB ID 2OMF) is shown as a blue line. The pore radii were determined using the program HOLE (28).

Figure 4

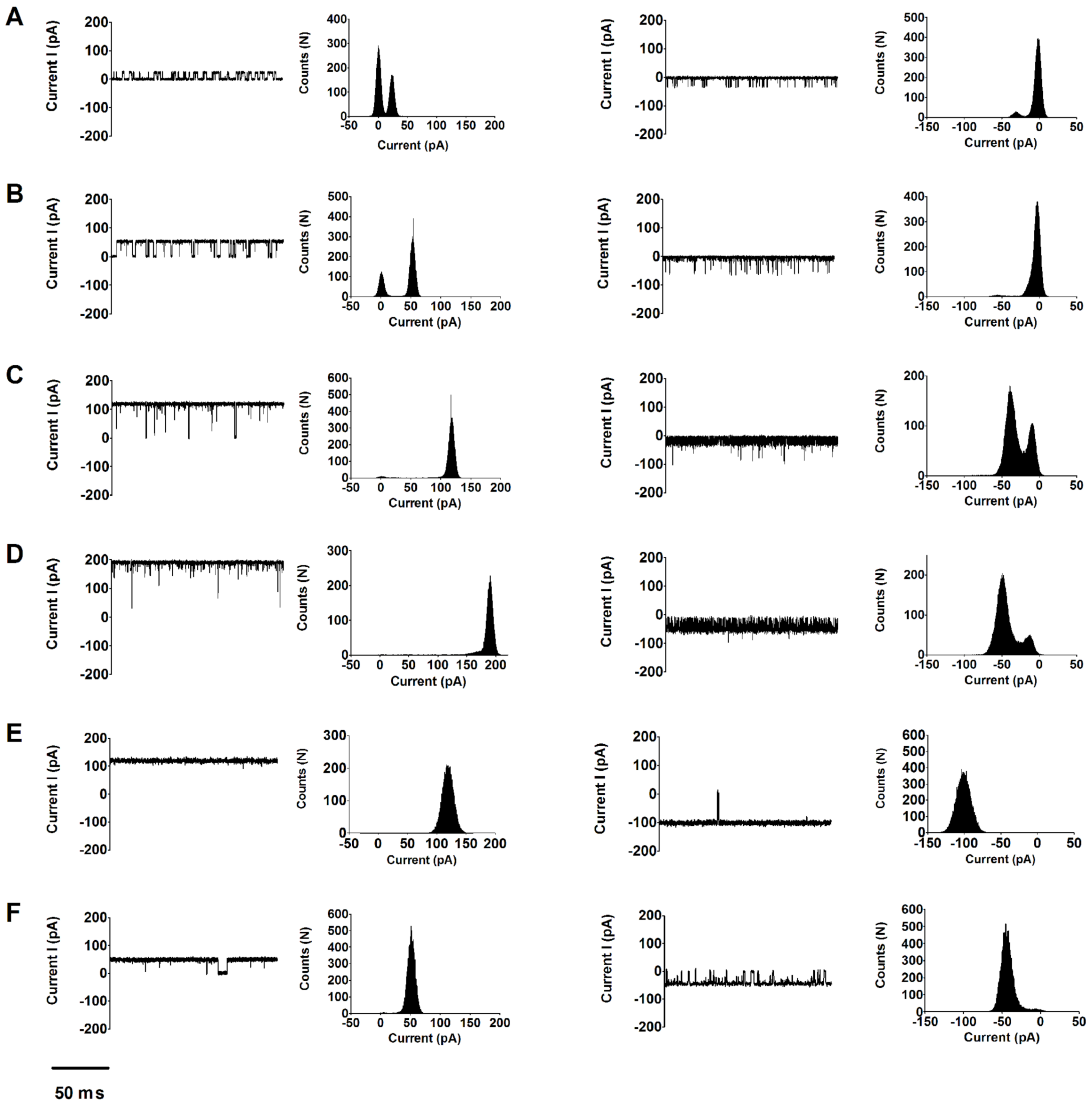

Figure 4 Single channel recordings of CymA show dynamics of the N-terminus. Traces are shown for the wild type channel at (A) +25 mV (left panel) and -25 mV (right panel), (B) +50 mV and -50 mV, (C) +100 mV and -100 mV and (D) +150 mV and -150 mV. Panel (E) shows conductance traces for ΔCymA at +100 mV (left) and -100 mV (right), whereas panel F shows traces for the R5A mutant protein at +50 mV (left) and -50 mV (right). The all-point histograms are shown in insets and represent quantitation of the different conductance states of the channel.

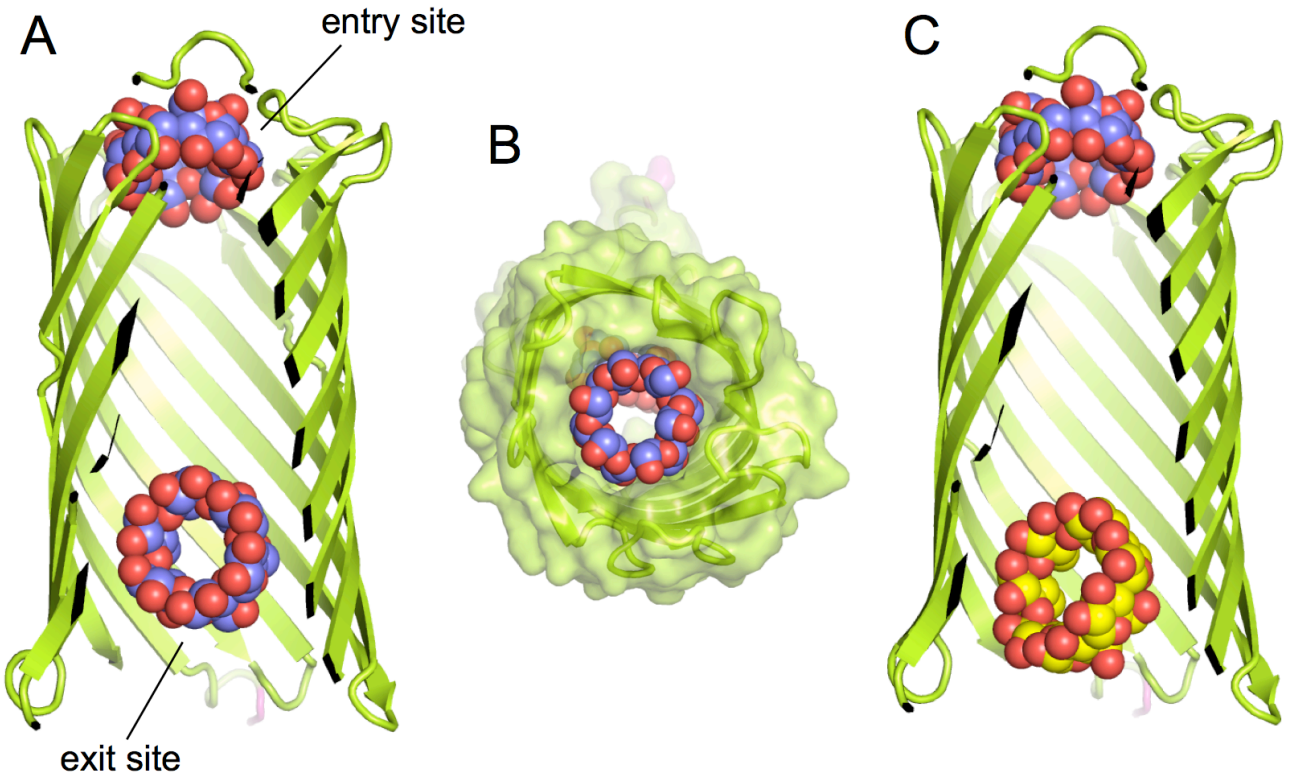

Figure 5 Cyclodextrins bind at two positions inside the CymA channel. (A) View of the interior of the CymA channel complexed with α-CD (PDB ID 4D5B), shown as a space-filling model with carbons coloured blue and oxygens red. The central panel (B) shows a view from the extracellular side, highlighting the similar dimensions of α-CD and the channel. The right panel (C) shows the co-crystal structure of CymA with β-CD (PDB ID 4D5B; carbons shown in yellow). In this structure, an α-CD molecule is bound in the entry site.

Figure 6

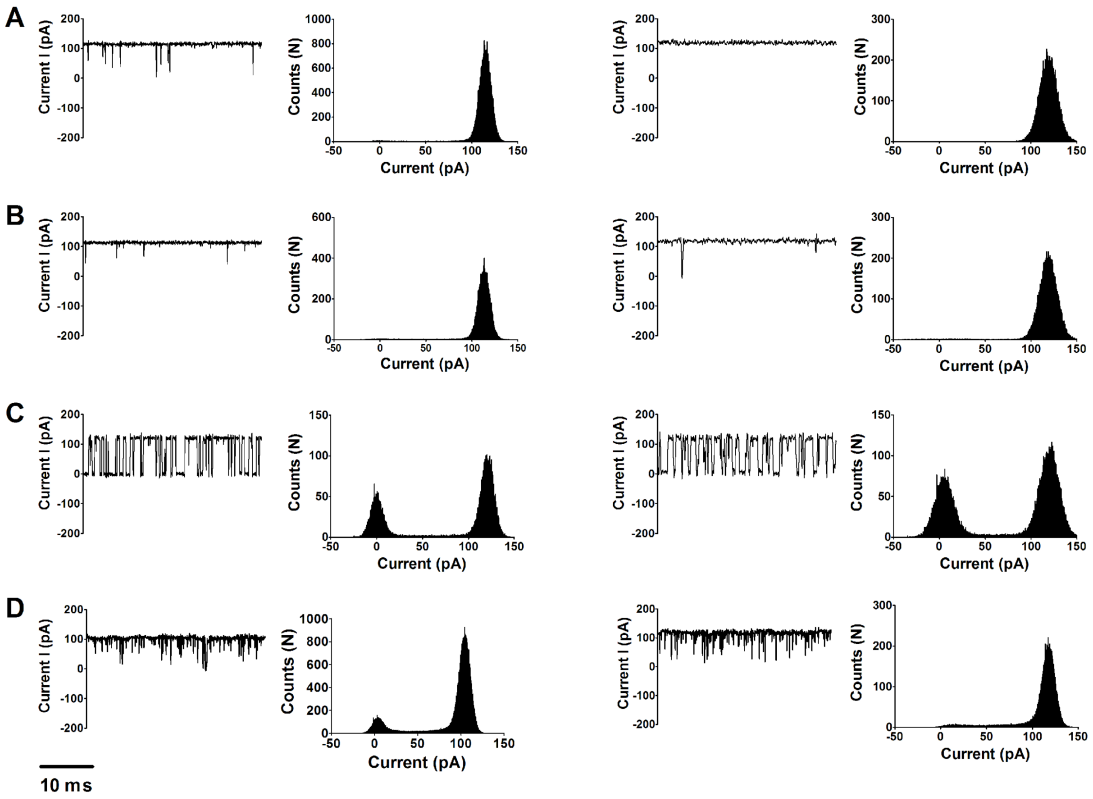

Figure 6 Directionality of CD binding to CymA probed by electrophysiology. Single channel conductance traces for wild type CymA (left panels) and ΔCymA (right panels) in the absence of α-CD (A) and in the presence of 10 μM α-CD added to the *cis* side (B) or to the *trans* side of the channel (C). Panel (D) shows traces obtained in the presence of 30 μM β-CD added to the *trans* side. All-point histograms are shown in insets and represent the different conductance states of the channel.

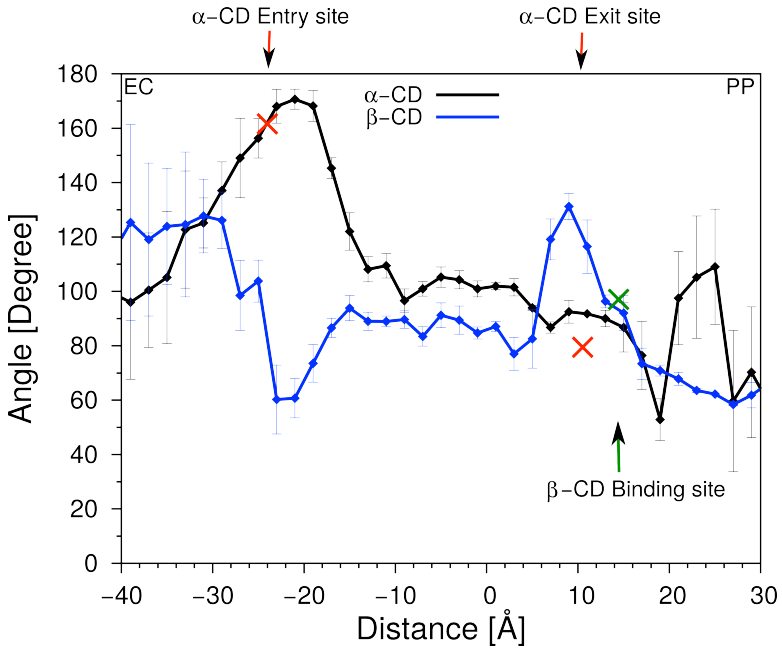

Figure 7 Orientation of α-CD and β-CD during translocation. Plots obtained from representative steered MD simulations showing the tilt angles of the CD molecules during transport through the CymA channel. The red and green crosses represent the orientations and positions of the α-CD and β-CD molecules in the respective crystal structures (PDB ID 4D5B and 4D5D).

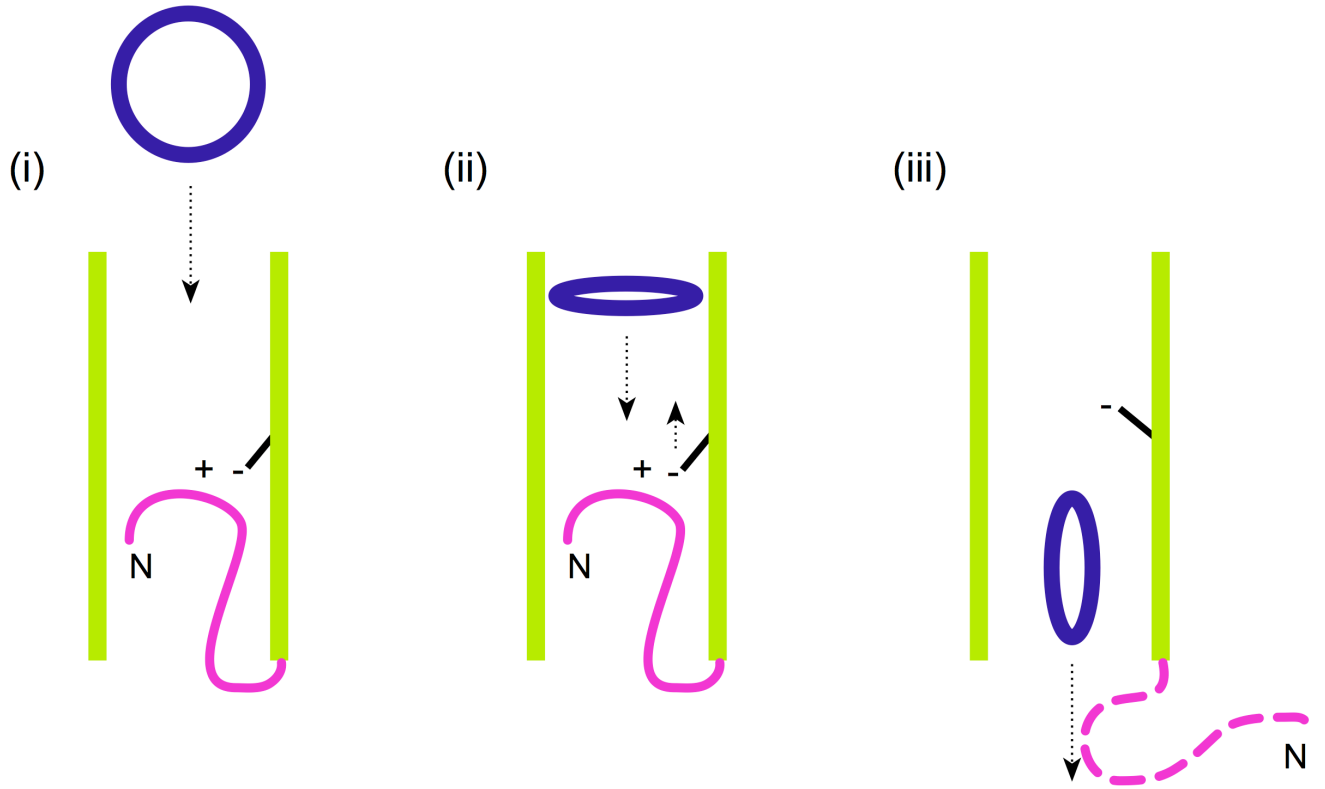

Figure 8 Schematic model for cyclodextrin (CD) transport by CymA. (i) The N-terminus of CymA (magenta) is located inside the barrel lumen. Arg5 in the N-terminus (+) interacts with carboxylate groups of glutamic acid residues (-) in the channel wall. (ii) After substrate capture by the entry site, the CD molecule rotates to lower its affinity for the channel and diffuses further. The N-terminus is expelled from the barrel by perturbation of the interactions of the N-terminus with the barrel wall. (iii) Upon expulsion of the N-terminus, the CD molecule diffuses into the periplasmic space.

SUPPLEMENTARY FIGURES

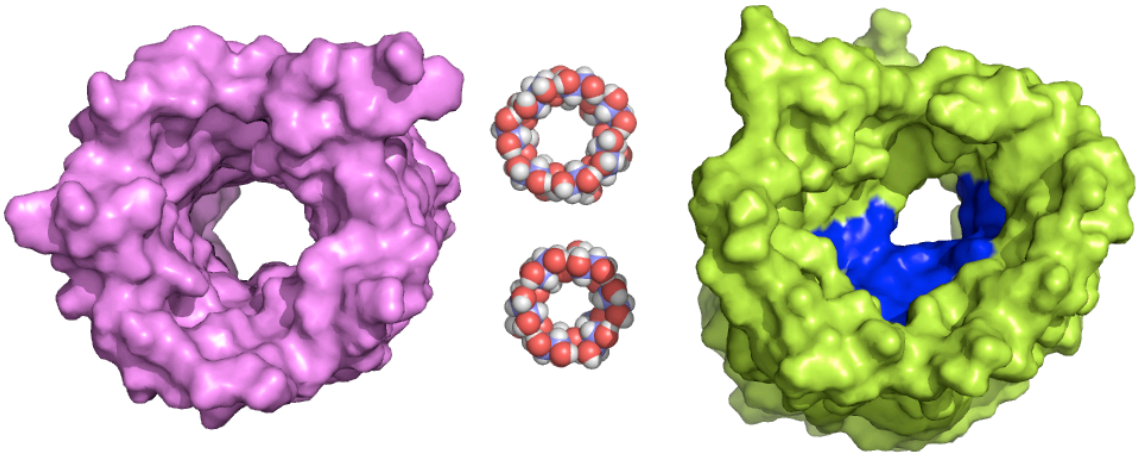

Fig S1 Relative dimensions of the channels of CymA (left) and *E. coli* OmpF compared to the CymA substrates α-CD and β-CD (top). Both substrates are clearly too large to pass through the OmpF channel. The channel-constricting loop L3 in OmpF is colored blue. Hydrogen atoms are included both for the channels and the substrates.

Fig. S2

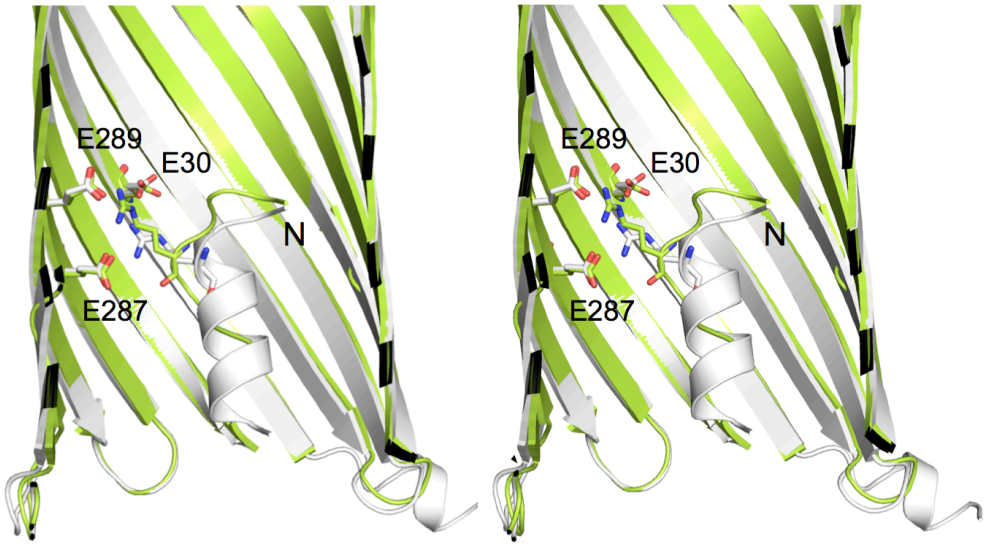

Fig. S2 Stereo diagram of cartoon superpositions for His-tagged SeMet CymA (white) and wild type CymA (green), highlighting the interactions of the N-terminal arginine residue (Arg5 in wild type CymA) with glutamic acid residues in the barrel wall.

Fig. S3

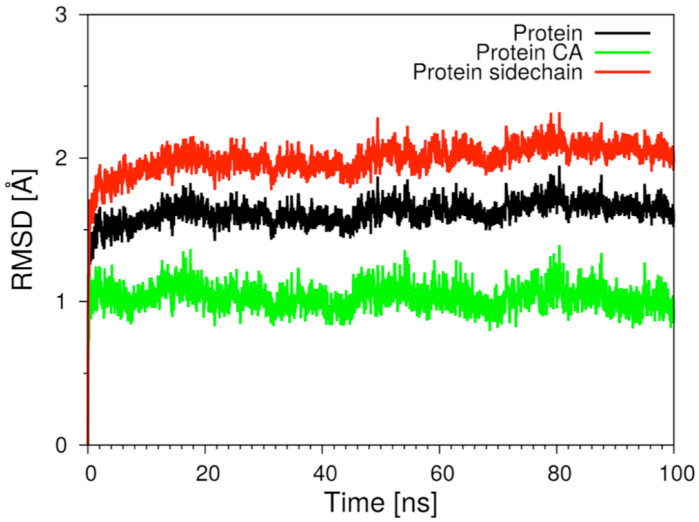

Fig. S3 Unbiased MD simulation of full-length CymA, showing that the structure is stable over the length of the simulation.

Fig. S4

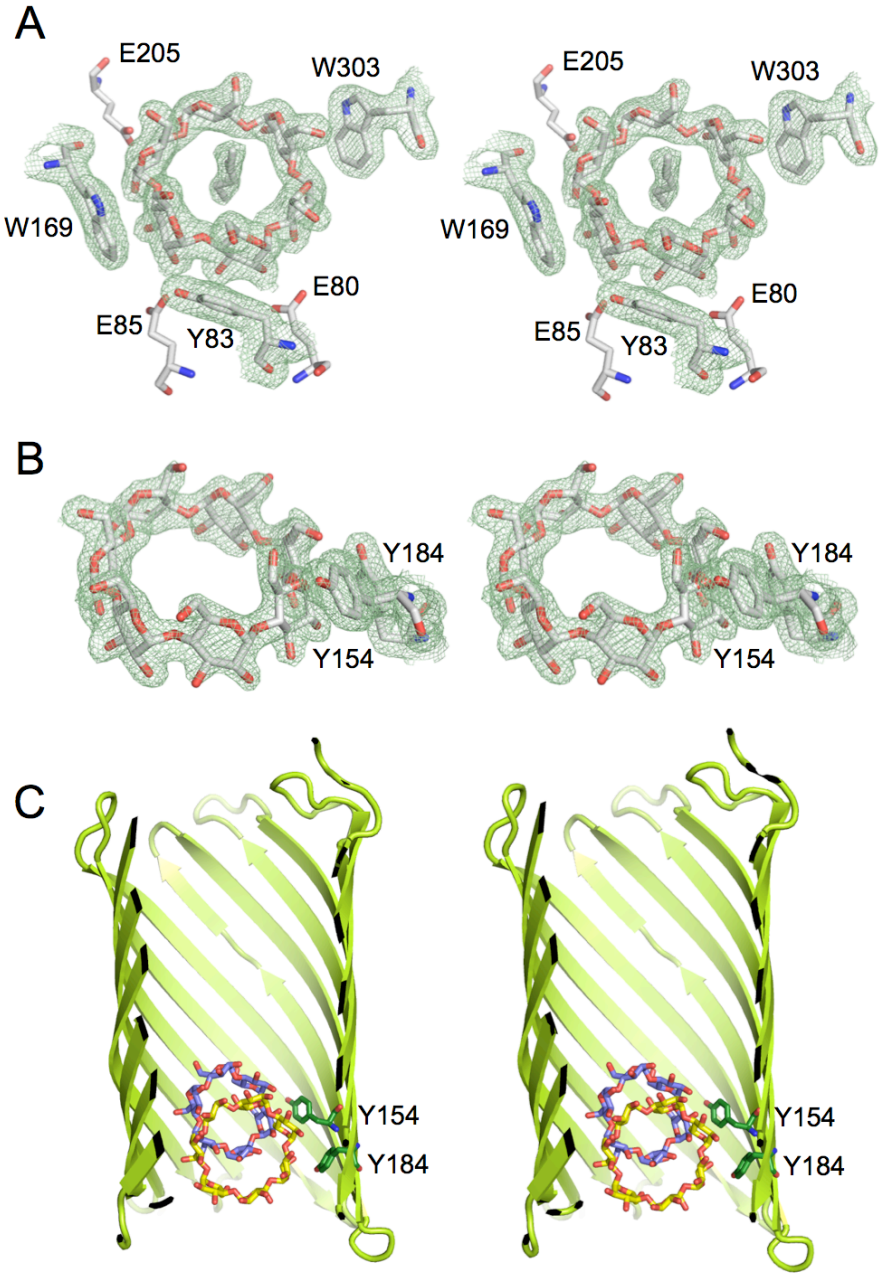

Fig. S4 CD binding sites in the complex of CymA and β-CD (PDB ID 4D5D). Stereo diagrams of $2F_o$-F_c electron density contoured at 1.0 σ for the entry site (A) and the exit site (B), showing the α-CD and β-CD molecule respectively. Aromatic residues making stacking interactions with sugars are shown, as well as the glutamic acid residues contacting the substrate within the entry site. Panel (C) shows a stereoview of a superposition of the β-CD co-crystal structure with that of α-CD (PDB ID 4D5B), showing the differences in binding between the exit site α-CD and the β-CD molecule.

Fig. S5

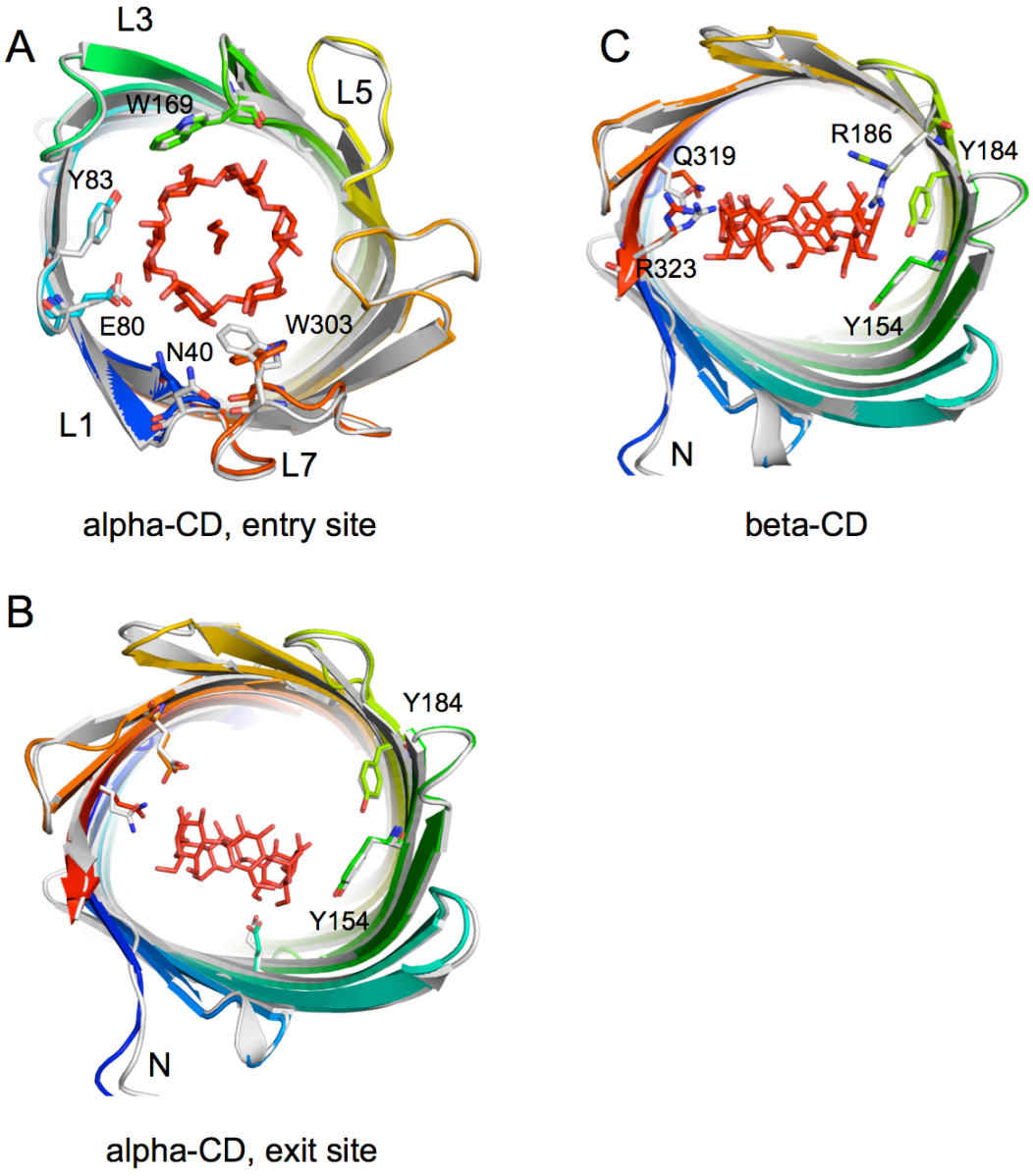

Fig S5 CD binding causes minor conformational changes in CymA. Views from the extracellular side (A) and from the periplasmic space (B,C) for the superposition of apo-CymA (PDB ID 4V3H; grey) and the complex of CymA and α-CD (PDB ID 4D5B; rainbow) and that of CymA and β-CD (PDB ID 4D5D; rainbow) . The aromatic residues Y154 and Y184 as well as the residues that change conformation (> 1 Å) upon CD binding are shown. The CD molecules are colored red.

Fig. S6

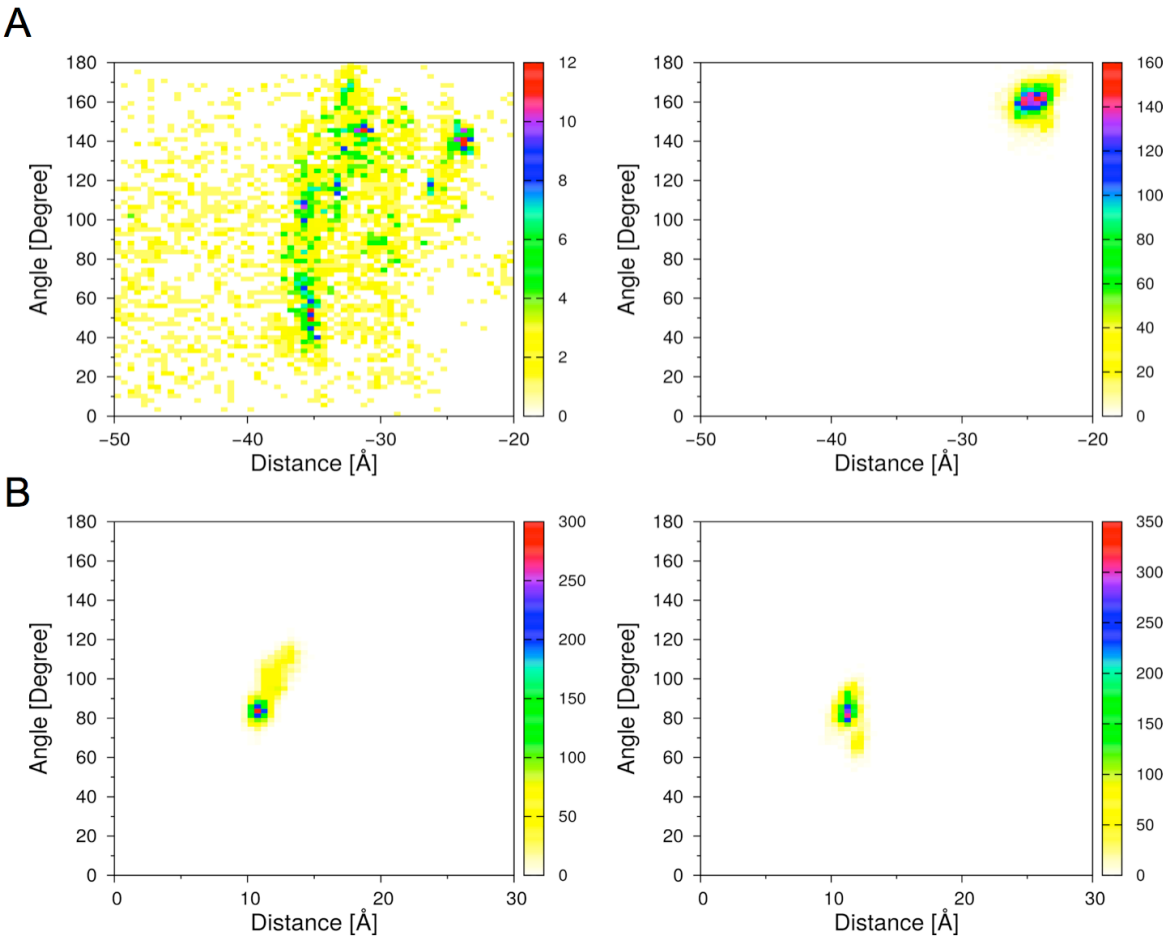

Fig. S6 Unbiased MD simulations (50 ns each, performed twice) for the α-CD molecules bound in the entry site (A) and exit site (B) of CymA (PDB ID 4D5B), showing the orientation (angle) of the substrate and its position within the channel. The simulations were performed with charged (left panels) and protonated (right panels) carboxylate groups contacting the substrate.

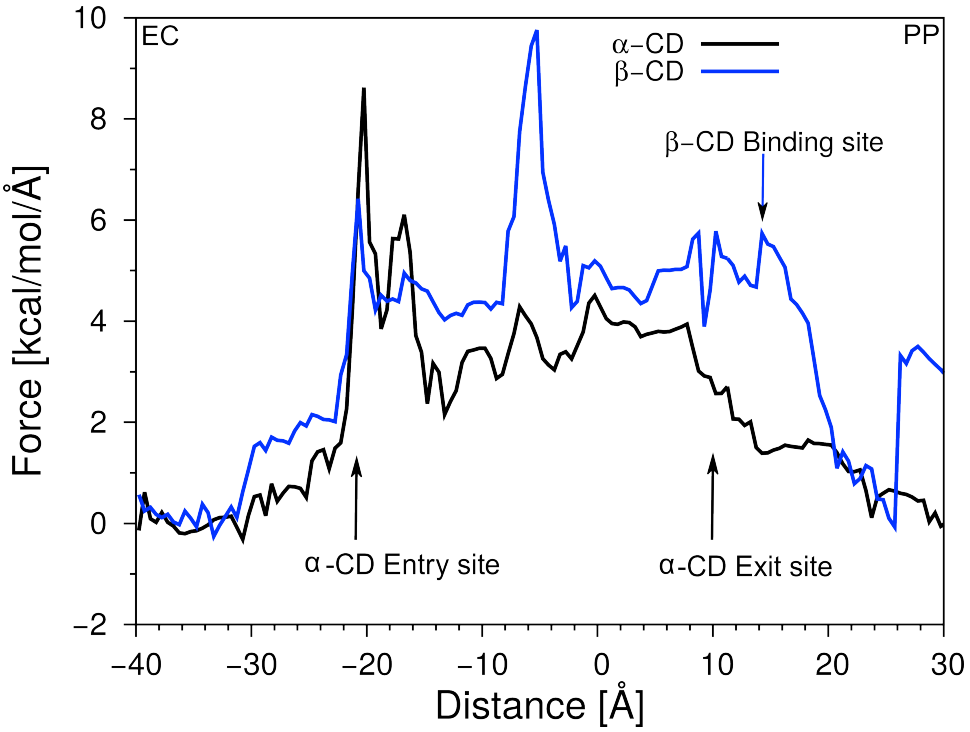

Fig. S7 Force profiles from constant-velocity steered MD simulations of α-CD and β-CD molecules along the CymA pore axis. Each CD molecule was placed on the extracellular (EC) side of the pore and pulled along the permeation pathway toward the periplasmic (PP) side of the pore. The profiles shown are averages of three simulations; two in the forward direction (EC to PP) and one in the reverse direction (PP to EC).

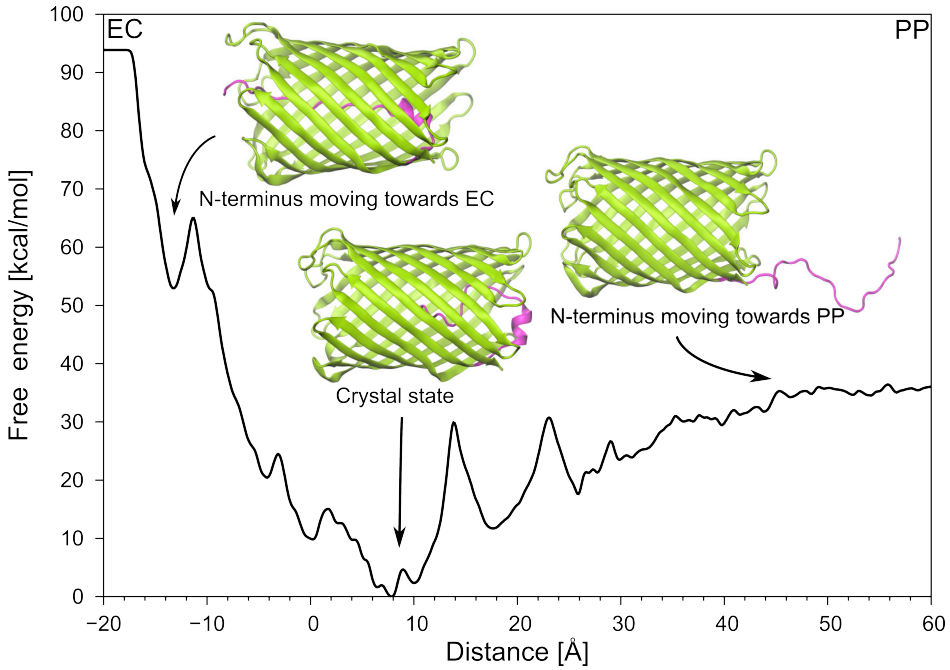

Fig. S8 One-dimensional free energy profiles (PMF) of the N-terminus along the channel axis determined using well-tempered metadynamics (14) implemented in PLUMED2 (29). The reaction coordinate distance is defined as the COM (center of mass) difference between the N-terminus (C_α atoms of residues 1-13) and the barrel (C_α atoms of residues 26 to 324) along the channel axis. Three representative snapshots of the protein are shown, including the crystallized state of the N-terminus corresponding to the lowest energy state.

Supplementary Movie Steered MD simulation of α-CD translocation through CymA